Can the integration hypothesis account for language evolution?*

Abstract

In a recent series of papers, Shigeru Miyagawa and colleagues propose an Integration Hypothesis for language evolution (Miyagawa 2016; Miyagawa et al. 2013, 2014; Nóbrega & Miyagawa 2015). This model suggests an 'emergent' view of language evolution in which two primitive systems, both occurring in animal communication, were integrated abruptly in our recent ancestors to form language as we know it. The proposed pre-existing systems are an 'E' system, for 'expressive', which occurs in learned birdsong, and an 'L' system, for 'lexical', which is said to have its origins in monkey alarm calls. What is distinctive about our hominin ancestors is that they alone attained the ability to 'merge' signals, so forming both words and syntax immediately via the integration of the L-system and the E-system. This view thus differs from much current work on the evolution of language, which suggests the gradual development of syntactic language from a pre-linguistic protolanguage. Here, I argue against the Integration Hypothesis, proposing that the two putative systems in animal communication do not have the requisite properties to form the basis for language. Conversely, the gradualist view of language evolution is well able to explain how grammatical and functional elements in language develop, uniquely in humans.

1. Outline of the Integration Hypothesis

In a number of recent papers, Shigeru Miyagawa and colleagues have proposed what is termed the Integration Hypothesis for language evolution (Miyagawa 2016; Miyagawa et al. 2013, 2014; Nóbrega & Miyagawa 2015). This hypothesis 'holds that language is an integration of two independently occurring systems in nature that underlie communication' (Miyagawa 2016), an event that is said to have occurred in recent evolutionary time, and in only one hominin species, *Homo sapiens*. This abrupt integration, according to Miyagawa, gives rise to the appearance of a rapid and recent emergence for language, though the two underlying systems themselves have an ancient history in animal communication. Miyagawa suggests, in fact, that language appeared within the past 100,000 years [100 kya], a view consistent with much contemporary work in the 'biolinguistic' framework (Chomsky 2010, 2012; Berwick & Chomsky 2011; Berwick 2011), but argued against by Dediu & Levinson (2013) and Tallerman (2014), among many others.

According to the Integration Hypothesis, language comprises two main components: sentences have an inner L-layer, for 'lexical', which provides basic lexical meaning, and an outer E-layer, for 'expressive', which has properties such as Tense, or Question, or Focus (Miyagawa et al. 2013, 2014). More precisely, Miyagawa (2016) proposes that the L-layer is composed of content words, while the E-layer consists of function words (i.e. grammatical elements). Nóbrega & Miyagawa (2015) suggest that in addition to phrases and sentences, simple words also comprise these two layers, so that all expressions display an inner L-structure, providing a semantic core, with an outer E-layer, the 'expression structure', which comprises contentless functional elements. These are categorial and inflectional features, providing information about the syntactic category of words, and also morphosyntactic information such as number, case or tense etc. In the case of words, the lexical layer is a ROOT which thus lacks syntactic category; it can only become, say, a noun, a verb or an adjective

when it combines via Merge with categorial and grammatical information, provided by the E-layer.

The two primitive systems said to form the basis for language are proposed to have antecedents in certain animal communication systems, both putatively co-occurring in our hominin ancestors as well. Miyagawa regards the Lexical system (L-system) as developing from the alarm calls of primates, and suggests that it is 'composed of isolated units of utterance that typically have a specific referent, such as "leopard," "snake," and "eagle"' (Miyagawa 2016), as occurring in vervets and numerous other present-day monkey species. The Expression system (E-system) of animal communication, on the other hand, has no 'vocabulary', but as Miyagawa sees it, is 'associated with birdsong, creat[ing] patterns without use of lexical items'. Here, the idea is that birdsong (and also, presumably, whalesong, which is at least as complex; Hurford 2012) has 'syntax without meaning' (Nóbrega & Miyagawa 2015), in the sense of displaying complex but meaningless patterns, sometimes with limited hierarchical structure. Early hominins are also supposed to have had an E-system, putatively in fact a song system that pre-dates language (Miyagawa 2016). When this E-system is integrated with lexical items from a pre-existing L-system in the course of human evolution, what results is full language. The crucial novelty in *H. sapiens* is proposed to be the sudden and recent appearance of the syntactic operation Merge, which creates hierarchical structures and integrates the two pre-existing systems, L and E (Nóbrega & Miyagawa 2015).

Note that in this model, there are presumed to be no earlier protolanguage stages in hominin evolution, contrary to much recent work in evolutionary linguistics (e.g. Bickerton (1990, 1995, 2009), Hurford (2012), Jackendoff (2002), Jackendoff & Wittenberg (2014)). As Nóbrega & Miyagawa (2015) observe, their 'emergent' view of language evolution 'assumes that the language faculty emerged late in historical development without any prior pre-syntactic stage'. Thus, the Integration Hypothesis explicitly denies a 'gradualist' view of language evolution, which portrays language as growing over the course of human evolution for perhaps half a million years or more, from a very primitive pre-syntactic stage when words did not combine, and gradually incorporating more linguistic features. However, Miyagawa (2016) does refer approvingly to the Darwinian idea of a pre-linguistic song stage, suggesting that 'the system underlying singing [plays] a critical role' in language evolution, and citing apparent homologies for speech in the brains of vocal-learning birds (Section 2.2).

In this short commentary I argue against the Integration Hypothesis on a number of grounds. A major problem is that the comparative animal evidence, if examined in detail, does not seem to provide either analogues or homologues to any of the highly specific properties of language; see also Hauser et al. (2014) for a detailed critique. This is the topic of section 2. Section 3 considers the relationship between full language, a putative protolanguage, and the Integration Hypothesis, and argues that a protolanguage system which gradually developed the properties of full language is entirely feasible, but does not stem from pre-existing systems in animal communication.

I turn first to an examination of the putative origins of the two systems in animal communication systems, and their relationship to the proposed L-system and E-system in human language.

2. The comparative animal evidence: are there homologues or analogues to the L-system and the E-system?

2.1 The L-system

Considering first the proposed origins of the L-layer, can we see anything wordlike in the calls of other primates? Nóbrega & Miyagawa's specific suggestion (2015) is that the typical monkey alarm calls are not actually analogues or homologues to words, but rather, to lexical roots or stems; they 'denote a conceptual content' but they lack the features of true words such as grammatical category and morphosyntactic properties. Nonetheless, there are a number of problems in proposing that there are similarities of any kind between primate calls and lexical roots. All these problems are extremely well known, and I rehearse them briefly here only to emphasize the fact that the Integration Hypothesis lacks empirical support. First, as Nóbrega & Miyagawa themselves note, primate calls are genetically transmitted, while lexical elements of any kind are culturally transmitted – in simple terms, entirely learned. This is absolutely not to deny a role for experience in sharpening up the appropriate usage of the various alarm calls by infant monkeys (Cheney & Seyfarth 1990). The point here is that the calls are developmentally canalized, which means that they are genetically coded, appearing without being learned, though in infant monkeys they are not immediately employed with the appropriate adult distribution. Thus, versions of the alarm calls start to occur in infants with too broad a usage at first, and are narrowed down and used correctly as the monkey matures. But this development is totally unlike that of lexical elements, which crucially require very precise environmental input to occur in the first place, since they are learned and not genetically coded.

Second, lexical elements are Saussurean signs (see Bouchard 2013); they are, critically, a learned linkage of two cognitive elements, namely concept and percept, and both elements are required in order to fulfil the criterion for signhood. Linguistic signs of course consist not merely of meanings, the *signifiés*, but of the *signifiants* as well, namely a mental image of an acoustic or gestural event: the sound-shape or manual shape of the word. The *signifiant*, then, is the internal representation of the sound or manual sign for the word in question, and this is always learned. This seems to be true whether we regard the word as the fundamental unit of language evolution, or the lexical root or stem; in both cases, the linkage is required. Primate calls cannot be Saussurean signs in this sense, since they lack *signifiants* precisely because their acoustic properties are <u>not</u> learned and are not linked by cultural convention to a meaning. This makes primate calls an unlikely candidate for the origins of the L-layer in language, even if the exact meanings of calls – or maybe strictly, their pragmatics/correct usages – are honed over time.

Third, it is specious to speculate about what primate alarm calls actually mean. We know fairly accurately the circumstances under which they are used, and we know their effects on the audience, but we cannot know much else at all about their 'meaning'. Nóbrega & Miyagawa (2015) describe the calls as 'systems that employ isolated uttered units that correlate with real-world references', while Miyagawa (2016) states that a simple way to view the call system is as 'a collection of vocal gestures that have a specific referent in the real world', suggesting 'leopard', 'eagle' and 'snake' to be some of the referents in question. This seems to overinterpret the available data: there is no evidence that, say, vervet monkeys have

a concept of the actual predators themselves, as opposed to, for instance, the specific direction from which the threat is coming, or even its manner (flight? Pouncing?). Alternatively, the alarm calls might be something akin to holistic instructions: 'Avoid danger by running to the thinnest branches', 'Scan the ground for danger' and so on. (There is some confusion here, since Miyagawa does note that the alarm calls are holistic, while nonetheless describing them as having specific referents.[1]) Nóbrega & Miyagawa (2015) suggest that monkey alarm calls constitute the 'closest approximation' to linguistic roots. In fact, calls seem rather more like propositions, as described above, than roots (Bickerton 2009), certainly in terms of their audience effects.

Fourth, neurological data does not support the idea of homology between primate calls and any linguistic elements. Miyagawa (2016) remarks on the 'existence of regions similar to Broca's and Wernicke's areas in the brains of some great apes'. As Wilkins (2012) notes, this is entirely predictable. Language has a 'mosaic nature'; it is not an organ, not localized in one specific brain region, and no brain region is exclusively devoted to language production or perception. It is thus fully expected that homologues to human brain regions critically involved in language will be found in the other apes, where they will have functions other than linguistic ones (as indeed both regions do in humans too). This expectation seems to be borne out for both Broca's and Wernicke's areas. However, there is much more to the neural structures underlying language than these two regions; for instance, Rilling et al. (2012), among others, argue for the importance of the arcuate fasciculus, a bundle of fibres which connects the 'traditional' language areas (Broca/Wernicke) in supporting syntax and lexical semantics. The connections in this region are far stronger in humans than in other primates, even the other great apes. It is known, too, that primate vocalization is supported by distinct brain regions than those handling language; see Arbib (2002), Ploog (2002), and the collection of papers in Scharff et al. (2013). Moreover, removing the Broca's area homologue in monkeys does not impair their ability to vocalize, and neocortical structures in the brain do not handle primate vocalization (Ploog 2002). Comparing the neural basis of animal calls and human speech, Rizzolatti & Arbib (1998:190) note that they are 'undoubtedly different phenomena', outlining in detail the 'marked difference in the anatomical structures underlying the two behaviors'. As Arbib (2012: 118) notes:

> Animals calls are mediated primarily by the cingulate cortex together with various subcortical structures as well as the brainstem, which bridges between the brain and the spinal cord. Speech (like language more generally) is mediated by a circuit whose main nodes include the classical Wernicke's and Broca's areas.

And Wilkins (2012) discusses an important factor in the structure of the brain: the lack of the POT (parietal-occipital-temporal junction) in other primates, even in the great apes most

[1] Miyagawa (2016) suggests that 'the lexical protolanguage view points to this L-system in our ancestors as forming the progenitors to language'. It is unclear to me whether this statement is intended to suggest that proponents of this view of protolanguage support the idea that words derive from primate calls; whatever the intention, I can state categorically that this is not the case. See, for instance, Bickerton (1990, 1995, 2009); Tallerman (2007).

closely related to humans. As Wilkins points out: 'The POT region in modern humans is essential not only to language but to basic conceptual structure. It accounts for the modality-neutral character of concepts' (2012:200). This is a rather crucial finding, since according to the Integration Hypothesis, it is primate calls that underlie the L-system of human language, and, which as noted above, are said to 'denote a conceptual content'. But if human-type concepts rely on the POT, which even apes lack, this strongly suggests that primate calls cannot be precursors of lexical elements in humans.

Fifth, if primate calls 'expanded [their] behaviour' to become root-like or word-like elements, as Nóbrega & Miyagawa (2015) suggest, then it is unexpected that humans retain a full panoply of primate calls, in fact 'gesture-calls', in the terminology of Burling (2005). These consist not only of cries of pain, anger and so on, but also laughter, sobbing, and non-vocal facial expressions, as well as postures and manual gestures that are thoroughly primate. In evolutionary terms, as Burling notes, the maintenance of our primate calls alongside their purported new linguistic usage would be very surprising if calls evolved into words – the wings of birds did not maintain either their prior form or function when they evolved from reptilian forelimbs.

These facts and others present a convincing case for a strict separation between primate calls and linguistic elements of any kind, rather than a homology. It seems, then, that if there is indeed an L-system in the alarm calls of modern primates, it is not the same system that underlies the human lexicon.

2.2 The E-system

The E-system in language comprises functional elements that lack independent status. In Miyagawa's terms (2016): 'The expression layer constitutes a chunk of the expression, typically a sentence, and its purpose is to give shape to the expression, such as interrogative, and any meaning that is associated with the entire expression, such as tense'. The Integration Hypothesis proposes that such a system has direct analogues in birdsong, which in some (though not all) species is characterized by having 'phonological syntax' (Marler 1998): in other words, regular combinatorial patterns without any semantic properties or semantic compositionality. There is, of course, an inherent contradiction in Miyagawa's work in this regard; as noted earlier, Nóbrega & Miyagawa (2015) characterize birdsong as displaying 'syntax without meaning', yet the E-system in language is intended to overlay meaning onto the expression, as the above quotation makes clear. How, then, did this meaning in the E-system arise in language?

Leaving this serious problem aside, we should consider what similarities there might be between song and language. Parallels of various kinds between learned birdsong and language are proposed in the literature (e.g Bolhuis et al. 2010), including the following: the appearance of a 'babbling' stage in both cases; the fact that juvenile birds of some species learn their song from a 'tutor', typically the father – thus fulfilling the traditional transmission criterion which also applies to language; and also the fact that this learning has a critical period ending around puberty in both humans and birds. As noted above, the Integration Hypothesis is even in part inspired by Darwin's thoughts on the parallels between birdsong and language, which led him to propose a pre-linguistic song stage in early humans – an idea

that Miyagawa regards as ‘particularly plausible’. Thus the Integration Hypothesis ‘incorporates Darwin’s idea that language began as song much like birdsong’, and Miyagawa ‘considers the system underlying singing to play a critical role’ in language evolution (Miyagawa 2016). One immediate problem with the proposed parallel is that once juvenile songbirds have learned their song from a male tutor during the ‘impressionable period’, in some species they do not themselves produce the memorized song until they reach sexual maturity, something that clearly does not occur in humans. For instance, the white-crowned sparrow learns its song in the critical period between 10 and 100 days old, but does not actually sing until it is one year old (Marler 1970). In the case of the northern cardinal, in which both male and female birds acquire learned song, Yamaguchi (2001) found that neither sex sings as juveniles during the early sensitive phase. The birds do use this period for learning and memorizing adult song, but only later do they practise the song extensively to perfect their own versions. Birdsong researchers in fact divide the behaviour of juvenile songbirds into two stages: in the sensory phase, the song of the tutor bird or birds is memorized but no singing occurs; in the sensorimotor phase, practice begins, leading to ‘sub-song’. For some species, such as zebra finches, the two stages do overlap, while in other species they evidently do not. Clearly, this latter kind of silent learning without any output is very unlike infants’ acquisition and use of language; in children, there is no silent phase and the emerging language becomes progressively more complex in well-defined stages as the child develops, and is used consistently throughout infancy and childhood.

Miyagawa and colleagues also go on to suggest a genetic song/language link, since recent research has discovered ‘molecularly analogous [brain] regions that are homologous for song and speech for vocal-learning birds and humans’ (Miyagawa 2016).[2] Nóbrega & Miyagawa (2015) consider this to ‘constitute strong genetic evidence for half of the Integration Hypothesis, namely the connection between birdsong and language’. Miyagawa’s later paper discusses this further, outlining two possibilities for this proposed connection: either the birdsong findings are simply a case of convergent evolution (with language), in which case we need not search further for a shared evolutionary history for learned song; or alternatively, the capacity for learned song is a true homology found in the avian and primate lineages and their common ancestor, and is thus several hundred million years old. However, in fact there is another possible explanation. It is now well known that right across animal phyla, genes are deeply conserved. For instance, though the complex eye has evolved convergently in numerous animal species that do not share an immediate common ancestor, its development is under the genetic control of the exact same Pax family of genes. Deep homology (Shubin et al. 1997) need not stem from shared phylogeny in the traditional sense, but rather, it results from the deeply conserved genetic building blocks for features that today may well be morphologically and phylogenetically very disparate (Shubin et al. 2009; McCune & Schimenti 2012). This might also be the case for the genes underlying vocal learning (Scharff & Petri 2011; Berwick et al. 2012), which would remove any support for a specific song/language link of the kind Miyagawa proposes. Presumably, vocal learning only actually surfaces in a lineage if the selection pressures make it adaptive.

[2] This sentence seems confused: are the brain regions analogous or homologous in birds and humans? A good survey of the brain pathways for vocal learning in birds and humans can be found in Jarvis (2004) and in the collection edited by Scharff et al. (2013).

Moreover, vocal learning not involving song at all has now been discovered in a few additional animal lineages, including cetaceans such as dolphins; bats; and more recently, elephants (Stoeger & Manger 2014). (Though see Hammerschmidt et al 2015 for evidence that mice, sometimes cited as exhibiting vocal learning, are in fact not in this category; the researchers found that both deaf infants and adult males lacking a cerebral cortex display normal calls and songs.) In sum, though the genes underlying the ability to learn vocally may be ancient, and thus perhaps found in vertebrates generally, the evidence suggests no necessary connection between birdsong and learned aspects of language; see Tallerman (2013) for a detailed critique of proposals for a putative song stage preceding language.

Leaving aside these biological questions, how robust is the proposed analogy between birdsong and the functional elements of language? As Miyagawa et al. (2013, 2014) note, the E-layer in birdsong has 'expressive' functions, such as marking territory or availability for mating. Clearly, it is not readily apparent that these functions are in any way similar to, say, the marking of interrogatives, or tense, or focalization in language, which are also the responsibility of the E-component: the expression layer in language 'communicates the intention of the speaker, such as the intention to ask a question' (Miyagawa 2016). What seems more problematic, however, is the fact that language signals these functions in highly specific ways that have no parallels in birdsong – most generally, via syntactic displacement. For instance, in English, polar questions are signalled by fronting an auxiliary to a position preceding the subject (*Are the kids leaving?*), while *wh*-questions are signalled by fronting of a phrase to a position to the left of the clause (*Which kids are leaving?*), as also in many other languages. Focalizations, cross-linguistically, are characterized by movement of the focalized phrase to a dedicated position in the clause; this may be the same position as that of *wh*-movement, or a distinct position, perhaps directly pre-verbal, as in Hungarian and Basque. Scrambling, as in Japanese and other languages with case-marked noun phrases, gives rise to non-canonical constituent orders amongst the arguments of a predicate, which alter the pragmatic intent of the speaker. Fundamentally, the very existence of syntax stems from the speaker's desire to signal expressive functions (in that, I concur with Miyagawa), but it does this in two ways that are highly specific to language, and which have no analogue in animal communication systems. Firstly, as in the above examples, syntax rearranges linear order within the clause, leading to permutations from the canonical order without changing the verb's valency. And secondly, syntax changes grammatical relations within the clause, for instance via passivization (*The cake was eaten __ by the children*), or via dative movement (*We gave a cake to the children > We gave the children a cake*) and so on, where the valency of the predicate is altered, in addition to permutations of word order.

Birdsong, in contrast, sometimes displays extensive variation, but the variations do not change the message in any way, as researchers in the field make clear: 'in general the different songs that a bird has convey exactly the same message as each other but do it in a highly varied manner' (Slater 2012: 97). Considerably stereotyped ordering sequences are also common in birdsong. For instance, though chaffinches have a quite complex song, involving different numbers of repetitions of a single syllable (i.e. a predictable sequence of one or more notes), nonetheless 'The sequence of syllable types within a song type is absolutely fixed' (Peter Slater, as cited by Hurford 2012:48). Hurford even suggests that fluctuations in the sequence of stereotyped song, for instance as occurring in the zebra finch,

may be the result not of deliberate variation, but instead may be a kind of performance error (2012: 47f). Whatever is the case, the crucial fact is that permutations in the combinatorial syntax of birdsong absolutely do not alter the message. Moreover, different song types may apparently number in the thousands in some species, such as the brown thrasher, but nonetheless there are no syllable combinations found within the same song; in other words, as Hurford notes (2012: 21), 'whatever syntax can be found in bird repertoires, they do not take advantage of its combinatorial possibilities'. It is thus rather difficult to see how song syntax can be said to be 'expressive', in Miyagawa's sense, when real diversity is limited and what does exist never produces meaningful variations.

If functional elements similar to those of language are lacking in birdsong and whalesong, are there any indications that our closer relatives, non-human primates, either exhibit such features naturally or else can learn them? This does not form part of Miyagawa's proposal, but it is certainly worth asking the question, since we would ultimately need to know how our ancestors (but only ours, and not those of other primates) came to possess an E-system. Therefore if we can find any antecedents for anything approaching an E-system in other primates, this would strengthen the case. In terms of whether an E-system can be attained by non-human primates under instruction, a number of attempts have been made to teach some aspects of language to great ape species in captivity, with varying degrees of success. In all cases, closed class or functional elements are conspicuously absent. The star pupil is certainly the bonobo Kanzi (Savage-Rumbaugh et al. 1998), who learned an impressively large set of lexigrams on a computerized keyboard, and was also adept at understanding English instructions, however strange (e.g. *Can you pour the ice water in the pottty? Go get the balloon that's in the microwave*). However, Kanzi cannot understand coordination (*Bring me a potato and a banana*), which seems to indicate an inability to understand the conjunction *and*, and he has problems understanding other grammatical elements such as article *a* or auxiliary *can*. In an overview of the 'linguistic' progress achieved by three species of language-trained animals, Kako (1999: 10) concludes that 'Kanzi shows no evidence of possessing anything like closed-class elements in his lexigrammatic system'. Nor does Kanzi modify the manual gestures he has learned in order to indicate functional elements, which is contrary to the methods used to signal grammatical categories in all sign languages.

In terms of what appears in the natural communication systems of primates, there are some suggestions in the recent literature that features rather akin to functional elements occur in the calls of Campbell's monkeys (e.g. Zuberbühler 2002, Ouattara et al. 2009a, b). When a 'boom' vocalization precedes an alarm call, it seems to attenuate the seriousness of the call, perhaps indicating that the threat is not imminent. Though Zuberbühler describes this behaviour as a 'syntactic rule', with the boom as a kind of modifier, it is notable that there is significant delay (c. 25 seconds) between the boom and the alarm call, indicating that these are not after all call combinations consisting of a functional plus a contentful element. And Ouattara et al. propose that these monkeys also produce a 'stem + suffix' sequence which indicates a broader meaning than the alarm call alone. For instance, *hok* is the eagle alarm call, but *hok-oo*, with an optional *-oo* unit, indicates 'general arboreal disturbance'.

To describe this element in monkey calls as a suffix is unproblematic providing it is understood that this has a merely metaphorical sense. Suffixes and other bound affixes in human language do not occur spontaneously or in isolation, but rather, have a life history that

typically sees them develop from adjoining free morphemes via well known pathways of grammaticalization and morphologization (e.g. Heine & Kuteva 2007; Carstairs-McCarthy 2010). Such attested development can be illustrated with examples of univerbation, a process in which a fixed phrase several words long turns into a single word. For instance, consider Latin phrases such as *clara mente* 'with a clear mind', reanalysed over time as a single word comprising a stem with a *-mente* suffix, which is the source of several productive derivational morphemes in Romance languages, such as the French adverbial suffix *-ment*, as in *exactement* 'exactly'. Or consider Germanic *baar*, a verb stem meaning 'carry', subsequently reanalysed as an adjective-deriving suffix on verb bases, as in modern German *fahr-bar* 'navigable' (< *fahren*, to go, travel); or for that matter, the newly productive English adjectival suffix *-oholic* or *-aholic*, as in *workaholic, chocoholic*, reanalysed from *alcohol-ic* (which also has its own derivational suffix). In fact, the entire life history of linguistic morphology – which arises over time and is lost, often in cycles – relies on the fact that language is compositional, so that words combine to form meaningful phrases. Adjoining words then often undergo various historical processes which change their status. Since semantic compositionality does not occur in primate calls, we cannot expect morphology either.

In sum, I have questioned Miyagawa's idea that the proposed E-system of language can realistically be said to be analogous to birdsong. There is also little evidence that the E-layer of language – the functional elements found in all languages – has a common phylogenetic history in either the productions or the cognition of primates.

One major question remains for the Integration Hypothesis. Where do the categorial and grammatical features that comprise the E-layer (Nóbrega & Miyagawa 2015) themselves come from in the evolution of language? This seems highly problematic. Certainly, there are no evident precedents in birdsong or whalesong, where each note, syllable or phrase has the same semantic status; there is no distinction between contentful elements and grammatical elements. The vital integration of systems in language evolution is supposed, at word level, to occur by combining roots (from the L-layer) with categorial and grammatical features from the E-layer. Yet these features appear to be total novelties in language, rather than reflecting some pre-existing aspect of animal communication.

A further point is crucially relevant. Miyagawa regards the E-layer and the L-layer as 'separate components, each with [their] own specific properties' (2016). However, as seen in the textbook examples of suffix formation sketched above, we actually know a great deal about how elements of the E-system in language come into being. Functional elements of all kinds, both free and bound morphemes, are historically – and putatively also in the course of language evolution – ultimately derived from content words, via the processes of grammaticalization and morphologization outlined by Heine & Kuteva (2002, 2007), Carstairs-McCarthy (2010) and Hurford (2012), among others. It is not, then, the case that the E-system of language arises independently of the L-system; rather, grammatical morphemes are in their entirety derived from lexical morphemes, often with intermediate stages in which grammatical elements themselves become further grammaticalized.

3. Language, protolanguage and the Integration Hypothesis

Nóbrega & Miyagawa (2015) deny the existence of a 'structureless, protolinguistic stage' in language evolution, and Miyagawa (2016) criticizes the idea of a protolanguage that would grow in complexity over time. However, it seems that the authors do in fact countenance a stage in which our hominin ancestors had a form of protolanguage containing what we might call proto-words: 'The pre-language stage is composed of root-like elements, each occurring in isolation of the others' (Nóbrega & Miyagawa 2015). These have reference, the authors note, but they have no syntactic category (N, V etc.), so cannot be combined and do not form hierarchical structures. In other words, at the root-only stage, there is no syntax. This view (with which I largely concur) seems startlingly similar to the very concept of a lexical protolanguage that Miyagawa and colleagues deny; see Bickerton (1990, 1995, 2009), Hurford (2012), Jackendoff (2002), Jackendoff & Wittenberg (2014), Tallerman (2007, 2014) and many others. Following such work, I also envisage a proto-word stage in early hominins, where root-like elements initially occurred singly; however, a subsequent development would be short sequences of proto-words, strung together without syntax and with no structure. Such sequences *are* combinations, but they are semantically- rather than syntactically-based, and the combinations are always optional, since there is no syntax to introduce obligatory dependencies and lexical requirements. Proto-words thus differ from words in full language by lacking syntactic category, grammatical features and subcategorization (such as a determiner requiring a following noun phrase, or a verb or preposition taking an object); any proto-word combinations that do occur are simply interpreted by pragmatic means at first.

This in fact seems largely consistent with Nóbrega & Miyagawa's (2015) claim that '"words" do not precede syntax in language evolution, but they are derived by such a system'. I support this statement, at least in the sense of acknowledging that words-as-we-know-them cannot be said to occur until they participate fully in syntactic structures. What I do not support is Miyagawa's view (2016) of how N and V came into being as syntactic categories. He asks:

> What could have shifted the isolated utterances of the L system into these abstract entities that can be combined into patterns furnished by the E system? It is not clear, but again, we can see what must have occurred: each item, as a process of abstraction, took on category, such as noun, verb, and so forth.

This seems to propose a rather mystical process, with no immediately obvious parallels in language emergence or development today – unlike the processes of grammaticalization and morphologization discussed in the previous section, which are still very much at work in modern language change. What is missing from Miyagawa's account is an indication of how such a process could have occurred, just by chance as it were, in early pre-language.

Instead of this random appropriation of category, I propose that proto-words first formed loose asyntactic combinations, interpreted using purely pragmatic principles such as Agent First, or Grouping, before they had obligatory (i.e. any actual) syntax (Jackendoff 2002; Hurford 2012). Subsequently, such principles often became syntacticized, so that they are part of the grammars of languages; Agent First, for instance, leads to the overwhelming prevalence of subject-initial constituent orders in languages today. If certain words are adjoined optionally, but on a regular basis, they typically start to acquire a syntactic

relationship, such as the head-dependent relationship, as we see in attested language change. A simple and well-known illustration of the power of frequent word combinations is the rise of the future marker *gonna* in English, which developed from a pure verb of motion, *go + to-V*, (e.g. *I am going to build a wall now*) into a grammaticalized auxiliary-like functional element with no necessary connection to motion at all (*I'm gonna sit here all day*; *He's gonna go away now* – note the need for lexical *go* in the second case). In such a manner, words gain full lexical properties, such as subcategorization. Such properties are akin to what have been termed 'edge features' in the Minimalist literature (Chomsky 2005, 2008; Boeckx 2011). As Boeckx (2011:53) puts it, 'lexical items are sticky'– they have features allowing them to be merged. Only once noun-like proto-words and verb-like proto-words gain distinct syntactic behaviours (such as being compulsory arguments of a predicate, or being a predicate that takes obligatory arguments) can we say that we have the two distinct categories of nouns and verbs: it is the syntactic contrast between them that gives rise to the two differentiated categories. These are, of course, just the same criteria that linguists use when starting to determine N vs. V categories in linguistic fieldwork today.

Though Miyagawa (2016) and I both envisage pre-words to have existed in our early ancestors, he argues against the notion that these could in any way form combinations before syntax existed. He states that one characteristic of language is that L-units do not combine directly, and in this respect they are like the monkey alarm calls that are said to be their progenitors. Instead, they require the intervention of the grammatical E-layer.[3] Miyagawa illustrates this with what he considers the present-day counterparts of L-units, which are pure lexical roots, unsupported by any functional grammar. In fact, though, the idea that L-units do not combine in modern language seems a rather Anglocentric view, and is not supported by data from a variety of languages other than English. For instance, Miyagawa (2016) notes that sequences such as those in (1), where the ungrammatical versions combine L-units directly, do not occur in English; the grammatical variants in parentheses show the intervention of elements from the functional E-layer, in bold type:

(1) a. *John book (i.e. John**'s** book)
b. *book long (i.e. The book **is** long)
c. *want eat pizza (i.e. want **to** eat pizza)

In other words, a possessive marker is required in (1a); a copula *be*, linking subject and predicate, in (1b); and an infinitival marker *to*, linking a verb with its complement, in (1c). Miyagawa states: 'What I contend is that anytime language combines elements, it must be done so [*sic*] with the intervention of the E system, and these morphological markings are evidence for this' (2016). Although the E-system is indeed ubiquitous in the world's languages, it is far from the case that all languages display the same kinds of grammatical elements in all constructions, or indeed display the same extent of usage of morphosyntax. Cross-linguistically, each of the L+L constructions in (1) in fact occurs

[3] Of course, it is technically conceivable that there might have been an earlier stage in which L-units – roots – could combine directly, as suggested above, but that following the appearance of the Merge operation, this became unavailable. I assume that this is highly unlikely.

frequently with no intervention from the E-system:

(2) a. llyfr Aled (Welsh)
book Aled
'Aled's book'

b. Ali marah (Malay)
Ali angry
'Ali is angry.'

c. Roedd o 'n [dymuno bwyta pizza]. (Welsh)
be.PAST.3SG he PROG want.INFIN eat.INFIN pizza
'He wanted to eat pizza.'

Example (2a) lacks an intervening possessive marker; (2b) lacks a copula; and the bracketed embedded clause in (2c) lacks an infinitival marker introducing the non-finite verb *bwyta.* Given that L-units can indeed combine directly in language today, it seems reasonable to suggest that such combination was also possible in a putative pre-language stage which only had lexical roots.

There is also plenty of evidence from attested language change that syntactic complexity arises spontaneously over time. It is well known that paratactic structures (e.g. two adjoining clauses without any dependency) give rise to hypotactic structures (Heine & Kuteva 2002, 2007), thus forming hierarchical structure (in this case, subordination) where previously only a flat structure existed. For instance, Heine & Kuteva (2007: 225f) note that the transition from paratactic clause combinations to hypotactic ones is clearly documented during the transition from Old Norse to Old Swedish, where, for instance, a demonstrative pronoun became a relative pronoun, thus forming clausal subordination; the history of English *that* is exactly parallel, with a demonstrative becoming a marker of subordination in clausal complements and relative clauses. Or consider the Amazonian language Canela-Krahô (Brazil), which nicely illustrates the same process (Heine & Kuteva 2002: 113f):

(3) a. rop ita
dog this
'this dog'

b. i-te [hūmre te rop curan ita] pupan
1-PAST man PAST dog kill REL see
'I saw the man who killed the dog.'

Whilst the underlined word is simply a demonstrative in (3a), it has taken on the role of a relative marker in (3b), thus forming a recursive structure.

We have seen, then, that languages themselves contain the seeds of their own development, with gradual accumulation of changes and syntactic features via the normal processes of grammaticalization and morphologization. There is no need to assume that a Merge operation is some sort of *deus ex machina* in language evolution, producing linguistic structure where

none previously existed.

4. Last words

In sum, plenty of data from attested language change indicates clearly the sources of grammatical features and functional vocabulary elements: languages themselves supply the raw materials for these elements. Conversely, the communication systems of animals do not seem to provide appropriate materials to be the progenitors of the E-system in language; in particular, birdsong is rather strikingly unlike the E-layer of language. I have also argued that the L-system does not develop from primate calls. It seems unlikely, on the basis of these arguments, that a new Merge operation simply integrates a pair of pre-existing systems in language evolution to form all the elements of language as we know it. I conclude that the Integration Hypothesis is unsupported.

*I would like to thank the reviewers for *JNL* for their very helpful and supportive comments. All errors are my own.

References

Arbib, M. A. (2002). The Mirror System, imitation, and the evolution of language. In Nehaniv, C. Chrystopher & Dautenhahn, K. (eds.), *Imitation in animals and artifacts*, Cambridge, MA: The MIT Press, 229–280.

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. New York: Oxford University Press.

Berwick, R. C. (2011). Syntax facit saltum redux: biolinguistics and the leap to syntax. In Di Sciullo & Boeckx (eds.), 65–99.

Berwick, R. C., Beckers, G., Okanoya, K. & Bolhuis, J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience* 4:5. doi: 10.3389/fnevo.2012.00005.

Berwick, R. C. & Chomsky, N. (2011). The biolinguistic program: the current state of its development. In Di Sciullo & Boeckx (eds.), 19–41.

Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.

Bickerton, D. (1995). *Language and human behavior*. Seattle: University of Washington Press.

Bickerton, D. (2009). *Adam's tongue: how humans made language, how language made humans*. New York: Hill & Wang.

Boeckx, C. (2011). Approaching parameters from below. In Di Sciullo & Boeckx (eds.), 205–221.

Bouchard, D. (2013). *The nature and origin of language*. Oxford: Oxford University Press.

Bolhuis, J. J., Okanoya, K. & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11: 747–759.

Burling, R. (2005). *The talking ape: how language evolved*. Oxford: Oxford University Press.

Carstairs-McCarthy, A. (2010). *The evolution of morphology*. Oxford: Oxford University Press.

Cheney, D. L. & Seyfarth, R. M. (1990). *How monkeys see the world: inside the mind of another species*. Chicago, IL: University of Chicago Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry* 36: 1–22.

Chomsky, N. (2008). On phases. In R. Freidin, C. Otero, & M. L. Zubizarreta (eds.), *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press, 133–166.

Chomsky, N. (2010). Some simple evo devo theses: how true might they be for language? In Larson, R. K, Déprez, V. & Yamakido, H. (eds.) *The evolution of human language: biolinguistic perspectives*. Cambridge: Cambridge University Press, 45–62.

Chomsky, N. (2012). *The science of language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Dediu, D. & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology* 4: 1–17.

Di Sciullo, A M. & Boeckx, C. (eds.) (2011). *The biolinguistic enterprise: New perspectives on the evolution and nature of the language faculty.* Oxford: Oxford University Press

Hammerschmidt, K., Whelan, G., Eichele, G. & Fischer, J. (2015). Mice lacking the cerebral cortex develop normal song: Insights into the foundations of vocal learning. Scientific Reports 5, article 8808, doi:10.1038/srep08808.

Hauser, M., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M., Watumull, J., Chomsky, N. & Lewontin, R. (2014). The mystery of language evolution. *Frontiers in Psychology* 5: 1–12. doi: 10.3389/fpsyg.2014.00401

Heine, B. & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Heine, B. & Kuteva, T. (2007). *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press.
Hurford, J. R. (2012). *The origins of grammar: language in the light of evolution*. Oxford: Oxford University Press.
Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar and evolution.* Oxford: Oxford University Press.
Jackendoff, R. & Wittenberg, E. (2014). What you can say without syntax: A hierarchy of grammatical complexity. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity*, (pp. 65–82). Oxford: Oxford University Press.
Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* 1016: 749–777.
Kako, E. (1999). Elements of syntax in the systems of three language-trained animals. *Animal Learning and Behavior* 27: 1–14.
McCune, A. R. & Schimenti, J. C. (2012). Using genetic networks and homology to understand the evolution of phenotypic traits. *Current Genomics* 13: 74–84.
Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology* 71: 1–25.
Marler P. (1998). Animal communication and human language. In Jablonski, N. & Aiello, L. (eds.). *The origin and diversification of language.* San Francisco, CA: California Academy of Sciences, 1–20.
Miyagawa, S. (2016). Integration hypothesis: a parallel model of language development in evolution. To appear in S. Watanabe, M. Hofman & T. Shimizu (eds.), *Evolution of the brain, cognition, and emotion in vertebrates*. Berlin: Springer.
Miyagawa, S., Berwick, R. C. & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology, 4*, 1–6.
Miyagawa, S., Ojima, S., Berwick, R. C. & Okanoya, K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in Psychology* 5: 1–6.
Nóbrega, V. A. & Miyagawa, S. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Frontiers in Psychology* 6: 1–8.
Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009a). Campbell's monkeys use affixation to alter call meaning. *PLOS One.* DOI: 10.1371/journal.pone.0007808
Ouattara, K., Lemasson, A. & Zuberbühler, K. (2009b). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *PNAS* 106: 22026–22031.
Ploog, D. (2002). Is the neural basis of vocalization different in non-human primates and *Homo sapiens*? In Crow. T. (ed.) *The speciation of modern Homo sapiens*. Oxford: Oxford University Press, 121–135.
Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, S. & Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience* 3: 11. DOI.ORG/10.3389/fnevo.2011.00011.
Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience* 21: 188–194.
Savage-Rumbaugh, S., Shanker, S. G. & Taylor, T. J. (1998). *Apes*, *language and the human*

mind. Oxford: Oxford University Press.
Scharff, C. & Petri, J. (2011). Evo- devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366: 2124–2140.
Scharff, C., Friederici, A. D. & Petrides, M. (eds.) (2013). Neurobiology of human language and its evolution: primate and nonprimate perspectives. *Frontiers in Evolutionary Neuroscience* DOI 10.3389/978-2-88919-111-6.
Shubin, N., Tabin, C. & Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature* 388: 639–648.
Shubin, N., Tabin, C. & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* 457: 818–823.
Slater, P. (2012). Bird song and language. In Tallerman & Gibson (eds.), 96–101
Stoeger, A. S & Manger, P. (2014). Vocal learning in elephants: neural bases and adaptive context. *Current Opinion in Neurobiology* 28: 101–107. DOI:10.1016/j.conb.2014.07.001
Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua* 117: 579–604.
Tallerman, M. (2013). Join the dots: A musical interlude in the evolution of language? *Journal of Linguistics* 49: 455–487.
Tallerman, M. (2014). No syntax saltation in language evolution. *Language Sciences* 46: 207–219.
Tallerman. M. & Gibson, K. R. (eds.) (2012). *The Oxford handbook of language evolution*. Oxford: Oxford University Press.
Wilkins, W. (2012). Towards an evolutionary biology of language through comparative neuroanatomy. In Tallerman & Gibson (eds.), 198–206.
Yamaguchi, A. (2001). Sex differences in vocal learning in birds. *Nature* 411(6835): 257–8
Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour* 63: 293–299.